however, may still retain significant flexibility. For this reason, many studies have attempted to introduce rigid, nonpeptide compounds which mimic the β-turn. Peptides with such nonpeptide β-turn mimic provide useful leads for drug discovery. Ball & Alewood (1990) *J Mol Recog* 3:55-64; WO 94/03494 (Kahn).

One of the revolutionary advances in drug discovery is the development of combinatorial libraries. Combinatorial libraries are a collection of different molecules, such as peptides, that can be made synthetically or recombinantly. Combinatorial peptide libraries contain peptides in which all amino acids have been incorporated randomly into certain or all positions of the peptide sequence. Such libraries have been generated and used in various ways to screen for peptide sequences which bind effectively to target molecules and to identify such sequences.

Many methods for generating peptide libraries have been developed and described. For example, members of the peptide library can be created by split-synthesis performed on a solid support such as polystyrene or polyacrylamide resin, as described by Lam *et al.* (1991) *Nature* 354:82 and PCT publication WO 92/00091. Another method disclosed by Geysen *et al.*, U.S. Pat. No. 4,833,092 involves the synthesis of peptides in a methodical and predetermined fashion, so that the placement of each library member peptide gives information concerning the synthetic structure of that peptide.

Considerable effort has been devoted to introducing structural constraints into combinatorial peptide libraries so that the member peptides represent more closely to their native counterparts. Houston *et al.* U.S. Pat. No. 5,824,483 describes a synthetic peptide library containing peptides featuring α-helical conformation and thus capable of forming coiled-coil dimers with each other. McBride *et al.* (1996) *J Mol Biol* 259:819-827 describe a synthetic library of cyclic peptides mimicking the anti-tryptic loop region of an identified proteinase inhibitor.

A complementary method for peptide library-based lead discovery is display of libraries on filamentous bacteriophages. This method allows the preparation of libraries as large as 10^{10}-10^{12} unique peptide members, many orders of magnitude larger than libraries that may be prepared synthetically. In addition to large library sizes, advantages of phage display include ease of library construction (Kunkel mutagenesis), coupling of the binding entity (displayed peptide) to a unique identifier (its DNA sequence), a selection protocol for amplifying rare binding clones in a pool, and the high fidelity of biosynthesis (compared to synthetic methods). Furthermore, rapid and inexpensive selection protocols are available for identifying those library members that bind to a target of interest. However, only natural peptides composed of L-amino acids may be displayed on phage, so the problem of defining three-dimensional structure-activity relationships is more difficult than it might be for a constrained peptidomimetic containing non-naturally occurring peptides or nonpeptide compounds. One possible solution to this problem is to use the structural constraints of a folded protein to present small variable peptide segments. Indeed, several small, stable proteins have been proposed as peptide display scaffolds. Nygren & Uhlen (1997) *Curr. Opin. Struct. Biol.* 7:463-469; Vita *et al.* (1998) *Biopolymers* 47:93-100; Vita *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96:13091-13096; Smith *et al.* (1998) *J. Mol. Biol.* 277:317-332; Christmann *et al.* (1999) *Protein Engng.* 12:797-806. Unfortunately, it is not clear that protein ligands obtained by this approach could be transformed to small-molecule drug leads. Epitope transfer from proteins to small peptides or to non-peptide small molecules remains an extremely challenging problem. Cochran (2000) *Chem. Biol.* 7:R85-R94.

Therefore, despite of extensive studies of the rules governing conformational preferences in natural peptides and the existence of several peptide library systems, those features necessary for structural stability of natural peptides remain poorly understood. In particular, there has been little systematic or quantitative assessment of the effect of residue substitutions and non-covalent interactions on structure.

Disclosure of Invention

The present invention provides a novel model system for assessing individual residue contributions to the stability of a defined peptide scaffold and for evaluating a series of substitutions presented in a combinatorial peptide library. The peptides of the invention are cyclized via disulfide bond between two cysteines within the peptide sequences. Amino acid substitutions at various defined residue sites influence the conformation of the cyclic peptides and their energy stabilities. The invention also provides methods of screening for and analyzing cyclic peptides with a specific secondary structure, β-turn, which provides further structural constraints to the peptides. The subject peptide library comprising a collection of β-turn bearing cyclic peptides can be used in screening for candidate biologically active molecules through molecular binding assays. Methods for such screenings are also provided by the instant invention. The compositions and methods of the invention can be used in analyzing the structure-activity relationships of peptides of interests, thereby providing insightful information for studies of molecular interactions involved in particular biological processes, as well as for rational design of therapeutic agents.

Brief Description of Drawings

Figure 1 depicts the design of bhp, a 10-amino acid model β-hairpin peptide. (A) Superimposed structures illustrate packing between disulfides and side chains of the closest non-hydrogen-bonded residues; (B) Schematic representation of the bhp model β-hairpin peptide with the side chains of the non-hydrogen-bonded residues 1, 3, 8 and 10 shown. X represents the varied residue selected from 19 of the 20 natural L-amino acids (excluding Cys).

Figure 2 shows the relative hairpin stability for substitution X in the bhp peptide sequence. (A) Cysteine effective concentrations (C_{eff}) relative to glutathione. Error bars are for ± one standard deviation; (B) Equilibrium free energy differences relative to the alanine peptide.

Figure 3 depicts NMR structure (minimized mean) of disulfide-cyclized hairpin bhpW. Sidechain W3 and the central turn residues G5 and N6 are shown in black. Sidechain L8 and the disulfide are shown in gray. Sidechains for the hydrogen-bonded residues (T2, E4, K7, T9) have been omitted for clarity.

Figure 4 (A-B) depicts NMR analysis of CD4 peptides. (A) Overlay of the fingerprint region of the COSY spectra for cd1 and cd2. (B) NMR structure ensemble for cd2 (20 models; two orthogonal views) shown superimposed on CD4 residues 37-46 from the crystal structure of gp120-bound CD4 (PDB entry 1GC1).

Figure 5 shows circular dichroism spectra of three peptide pairs of Example 2.

possibly, other as yet poorly understood sequences. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

The term "coat protein" means a protein, at least a portion of which is present on the surface of the virus particle. From a functional perspective, a coat protein is any protein which associates with a virus particle during the viral assembly process in a host cell, and remains associated with the assembled virus until it infects another host cell. The coat protein may be the major coat protein or may be a minor coat protein. A "major" coat protein is a coat protein which is present in the viral coat at 10 copies of the protein or more. A major coat protein may be present in tens, hundreds or even thousands of copies per virion.

The terms "electroporation" and "electroporating" mean a process in which foreign matter (protein, nucleic acid, etc.) is introduced into a cell by applying a voltage to the cell under conditions sufficient to allow uptake of the foreign matter into the cell. The foreign matter is typically DNA.

A "fusion protein" is a polypeptide having two portions covalently linked together, where each of the portions is a polypeptide having a different property. The property may be a biological property, such as activity in vitro or in vivo. The property may also be a simple chemical or physical property, such as binding to a target molecule, catalysis of a reaction, etc. The two portions may be linked directly by a single peptide bond or through a peptide linker containing one or more amino acid residues. Generally, the two portions and the linker will be in reading frame with each other.

"Heterologous DNA" is any DNA that is introduced into a host cell. The DNA may be derived from a variety of sources including genomic DNA, cDNA, synthetic DNA and fusions or combinations of these. The DNA may include DNA from the same cell or cell type as the host or recipient cell or DNA from a different cell type, for example, from a mammal or plant. The DNA may, optionally, include selection genes, for example, antibiotic resistance genes, temperature resistance genes, etc.

"Ligation" is the process of forming phosphodiester bonds between two nucleic acid fragments. For ligation of the two fragments, the ends of the fragments must be compatible with each other. In some cases, the ends will be directly compatible after endonuclease digestion. However, it may be necessary first to convert the staggered ends commonly produced after endonuclease digestion to blunt ends to make them compatible for ligation. For blunting the ends, the DNA may be treated in a suitable buffer for at least 15 minutes at 15°C with about 10 units of the Klenow fragment of DNA polymerase I or T4 DNA polymerase in the presence of the four deoxyribonucleotide triphosphates. The DNA may then purified by phenol-chloroform extraction and ethanol precipitation. The DNA fragments that are to be ligated together are put in solution in about equimolar amounts. The solution will generally also contain ATP, ligase buffer, and a ligase such as T4 DNA ligase at about 10 units per 0.5 μg of DNA. If the DNA is to be ligated into a vector, the vector is first linearized by digestion with the appropriate restriction endonuclease(s). The linearized fragment is then treated with bacterial alkaline phosphatase or calf intestinal phosphatase to prevent self-ligation during the ligation step.

A "mutation" is a deletion, insertion, or substitution of a nucleotide(s) relative to a reference nucleotide sequence, such as a wild type sequence.

A2 and A4 are selected from the group consisting of amino acids W, Y, F, L, M, I, and V;

A3 is any naturally occurring L-amino acid and n is an integer that is selected from the group consisting of 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12; and

C1 and C2 are joined together by a disulfide bond thereby forming a cyclic peptide.

In one preferred embodiment, the peptides of the invention have a β-branched residue having two non-hydrogen substituents on the β-carbon of the amino acid residue at position A1 or A5 or both. More preferably, A1 or A5 is threonine (T). Even more preferably, both A1 and A5 are threonine residues.

According to another preferred embodiment, the peptides have an aromatic residue W, Y, F or H at position A1 or A5 or both. More preferably, A1 or A5 is W. Additional preferred peptides of the invention have a branched aliphatic residue I, V or T at A1, A5 or both.

In another preferred embodiment, the peptides of the invention have an aromatic residue W, Y or F at position A2 or A4 or both. More preferably, A2 or A4 is W; and even more preferably, A2 and A4 are Ws. Another preferred embodiment include peptides having an unbranched aliphatic residue L or M at position A2 or A4 or both; more preferably A2 or A4 is Leucine. Still other preferred peptides have a branched aliphatic residue I or V at position A2 or A4 or both.

In the peptides of the invention, the number of the A3 residues n can be 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12; preferably 4, 5, 6, 7, 8, 9, or 10; and more preferably 4, 5 or 6. In one embodiment, n is 4 and the resulting peptides are decamers. In these decamers, the residue sites A1, A2, A4 and A5 are each from a selected group of amino acid residues as described above, whereas the middle $(A3)_4$ is a tetrapeptide sequence with varying amino acids. In one aspect of the invention, the $(A3)_4$ tetrapeptide sequence is selected from those favorable to forming a β-turn structure, including but not limited to EGNK, ENGK, QGSF, VWQL and GPLT.

In one aspect, the library of the instant invention contains at least about 10^2 member peptides, each of which has at least one amino acid variation from others. Preferably, the library contains at least about 10^4 peptides, more preferably about 10^{10} peptides and even more preferably at least about 10^{12} peptides. According to various embodiments, the amino acid variation occurs at defined positions within the peptides. For example, variations can occur at non hydrogen-bonded (NHB) strand sites (e.g., A1/A5) or hydrogen-bonded strand sites (e.g., A2/A4); a residue and its cross-strand counterpart (e.g., A1/A5 or A2/A4) can have same or different amino acids. Variations can also occur at the middle $(A3)_n$ sites, wherein A3 can be any of the 20 naturally occurring L-amino acids.

The carboxy terminal end and the amino terminal end of the cyclic peptide may be protected with any known protecting groups or may be bonded to other amino acid residues (generally naturally occurring residues), both in the (L) and in the (D) form through conventional amide peptide bonds. The protecting groups and additional residues can be added using conventional peptide synthesis techniques. Generally from 1 to about 50, preferably from 1 to about 20, amino acid residues may be present on each of the carboxy and amino terminal positions, independently. These additional residues may be part of a known protein containing a beta turn of interest or may be any other desired sequence of residues. These additional residues may be added to determine

(IYSNPDGTWT) was compared to a second peptide with a different turn sequence (IYSNSDGTWT). Both peptides were estimated by NMR as 30% hairpin in water (de Alba *et al.* (1996) *Fold. Des.* 1:133-144). Further variation of this peptide, predominantly in the turn sequence, yielded hairpins of various structures and mixed populations. Generally no one conformer population exceeded 50% (de Alba *et al.* (1997) *J. Am. Chem. Soc.* 119:175-183). In a final study, the three N-terminal residues in peptide ITSNSDGTWT were replaced with various sequences. Again, mixed conformers were frequently observed and populations of a given hairpin conformer were generally less than 50%: one peptide (YITNSDGTWT) did form a register-shifted hairpin that was highly populated (80%; de Alba *et al.* (1997) *Protein Sci.* 6:2548-2560). The authors of these studies conclude that conformational preferences of the turn residues dominate cross-strand interactions in determining the stability of hairpins, at least in these short model peptides.

Analysis of hairpin sequences in crystal structures has allowed the design of a different series of β-hairpin peptides. The target structure was a type I' turn flanked by three-residue strands. Arg-gly sequences were added to the ends to improve solubility. The peptide RGITVNGKTYGR is partially folded into a hairpin conformation (about 30%) as determined by NMR (Ramirez-Alvarado *et al.* (1996) *Nat. Struct. Biol.* 3:604-612). The importance of strand residues is indicated by replacement of the ile and val, the lys and tyr, or all four residues with alanine. None of the alanine-substituted peptides showed any tendency to form a hairpin. The same authors reported a second series of experiments in which position i+1 of the turn was varied (asn to asp, ala, gly or ser). No peptide was more structured than the original sequence with asn in the turn (Ramirez-Alvarado *et al.* (1997) *J. Mol. Biol.* 273:898-912). A review describing this work stated that adding glu-lys pairs to the termini of the model peptide stabilized the hairpin but did not give further details (Ramirez-Alvarado *et al.* (1999) Bioorg. Med. Chem. 7:93-103).

Another model peptide series (RYVEVXGOrnKILQ) has yielded evidence for hairpin formation in water. Residue X as D-pro or L-asn yields characteristic NOEs and alpha-H shifts, but the L-pro peptide is unfolded. No population estimates are given, but D-pro appears to give the more stable hairpin (Stanger & Gellman (1998) *J. Am. Chem. Soc.* 120:4236-4237).

A designed 16-residue peptide (KKYTVSINGKKITVSI) based on the met repressor DNA binding region formed a hairpin structure in water with an estimated population of 50% at 303 K. Truncation of one strand showed that the turn was populated without the strand interactions, although to a lesser degree (35%). An analysis of the thermodynamic parameters for hairpin formation showed that folding is enthalpically unfavored and entropically driven, with $\Delta G = -0.08$ kcal/mol at 298 K (Maynard & Searle (1997) *Chem. Commun.* 1297-1298; Griffiths-Jones *et al.* (1998) *Chem. Commun.* 789-790; Maynard *et al.* (1998) *J. Am. Chem. Soc.* 120:1996-2007).

A final hairpin peptide (GEWTYDDATKTFTVTE) derived from the B1 domain of protein G (GB1) has some features relevant to the peptides of the invention. Unlike the above described model hairpins, the GB1 hairpin has four threonine residues at hydrogen-bonded sites in the strands, including one thr-thr cross-strand pair. This is generally believed to be an unfavorable pairing. In addition, there are trp-val and tyr-phe pairs at adjacent nonhydrogen-bonded sites that might interact to form a small hydrophobic core. The reported data indicate that the GB1 peptide formed a well-populated hairpin (about 50%) in water. The data are consistent

The sequence of positive clones is identified by sequencing the encoding DNA. Peptides corresponding to the displayed sequences (i.e., 12-mers) are synthesized using standard solid-phase methods. The peptides are then assayed using an appropriate biological or binding assay to determine their potency. Peptides can be evaluated for hairpin structure using any of the known techniques outlined above: circular dichroism, NMR, or disulfide equilibrium. Substitutions may then be made in the peptides to determine the relative contributions of the selected turn residues to binding. Ideally, these substitutions will not disrupt the scaffold structure. Once the nature of the binding motif is understood, the turn sequence can then be transferred onto a suitable organic scaffold for further optimization.

While the invention has necessarily been described in conjunction with preferred embodiments, one of ordinary skill, after reading the foregoing specification, will be able to effect various changes, substitutions of equivalents, and alterations to the subject matter set forth herein, without departing from the spirit and scope thereof. Hence, the invention can be practiced in ways other than those specifically described herein.